DOCUMENT RESUME

ED 055 097 TM 000 813

AUTHOR Aleamoni, Lawrence M.

TITLE MERMAC: A Model and System for Test and Questionnaire Analysis.

INSTITUTION Illinois Univ., Urbana. Office of Instructional Resources.

PUB DATE Mar 70

NOTE 14p.

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Computer Oriented Programs; Course Evaluation; Course Objectives; Data Analysis; *Effective Teaching; *Feedback; *Formative Evaluation; *Instructional Improvement; Item Analysis; Learning; Models; Questionnaires; Scores; Student Attitudes; Test Construction

IDENTIFIERS *MERMAC System

ABSTRACT

An evaluation model was developed to aid the classroom instructor in more effectively teaching his course by providing him with the basic data upon which to make judgements about student abilities, course attitudes, grade distributions, test appropriateness, etc. The MERMAC system was designed so that each instructor could obtain the data in a usable format and also be able to provide relatively immediate feedback to the students taking an examination or filling out a course evaluation questionnaire. A description of the software system and its interface to the evaluation model is presented. (Author)

research report # 33 0

research report

MERMAC: A Model and System for Test and Questionnaire Analysis*

Lawrence M. Aleamoni

Measurement and Research Division
Office of Instructional Resources
University of Illinois
507 East Daniel Street, Champaign

March, 1970

measurement and research division
office of instructional resources
university of illinois
urbana, illinois

Abstract

An evaluation model was developed to aid the classroom instructor in more effectively teaching his course by providing him with the basic data upon which to make judgements about student abilities, course attitudes, grade distributions, test appropriateness, etc. The MERMAC system was designed so that each instructor could obtain the data in a usable format and also be able to provide relatively immediate feedback to the students taking an examination or filling out a course evaluation questionnaire. A description of the software system and its interface to the evaluation model is presented.

MERMAC: A Model and System for Test and Questionnaire Analysis[1]

Lawrence M. Aleamoni

We know that what is taught in the classroom may vary from excellent to poor, applicable to useless, and understood to misunderstood, and that instruction exists if you have an instructor, a textbook, a TV presentation, or a film presentation. What is LEARNED, however, is not necessarily that which is taught. Therefore, in order to determine the extent, type, and degree of learning in a classroom, some method of evaluation is necessary.

Let us begin with the premise that if instruction is to be effective, then there must be some interaction between what the instructor presents and what the students comprehend. Therefore, in order to aid the instructor in improving his instruction, a course evaluator must concentrate on one major aspect of the instructor-student interaction, which is the knowledge of student learning. This knowledge should be supplied to (a) the student so that he may be constantly aware of his progress in comparison to others in his class and in comparison to course standards and (b) the instructor so that he may alter his instructional rate, content, or method in order to suit the observed learning.

What the student generally studies and learns tends to be the material that is tested, not necessarily the material that is presented. Tests used in a course represent, in a very practical manner, the direction an instructor thinks the students should go which, in turn, determines the

---

[1]The manual and system can be obtained by writing the University Press, University of Illinois, Urbana, Illinois 61801.

emphasis that the students will place on the material treated by the test. Good tests can, therefore, orient student learning, while the development of the tests and the knowledge of student responses can help to improve instruction.

Description

The bases, therefore, for developing a model and system for test analysis should be to:

1. assist the instructor in developing valid and reliable tests; and
2. provide rapid and meaningful feedback to the instructor and students.

MERMAC was developed with the above bases in mind. It is made up of two sets of programs (see Figure 1): (a) utility (data manipulation) programs, and (b) test and questionnaire analysis programs. The seven utility programs allow the user to copy, edit, match, merge, sequence, sort, and recode the input data. Generally, the purpose of these programs is to prepare the data for input to the test and questionnaire analysis programs. The six test and questionnaire programs allow the user to:

1. Score item data and produce up to forty subscores for each individual. Each item and response may be weighted to arrive at the scores. Any item may be included in more than one subscore and be weighted differently in each. An example of the SCØRE program output is presented in Figure 2.

MERMAC System

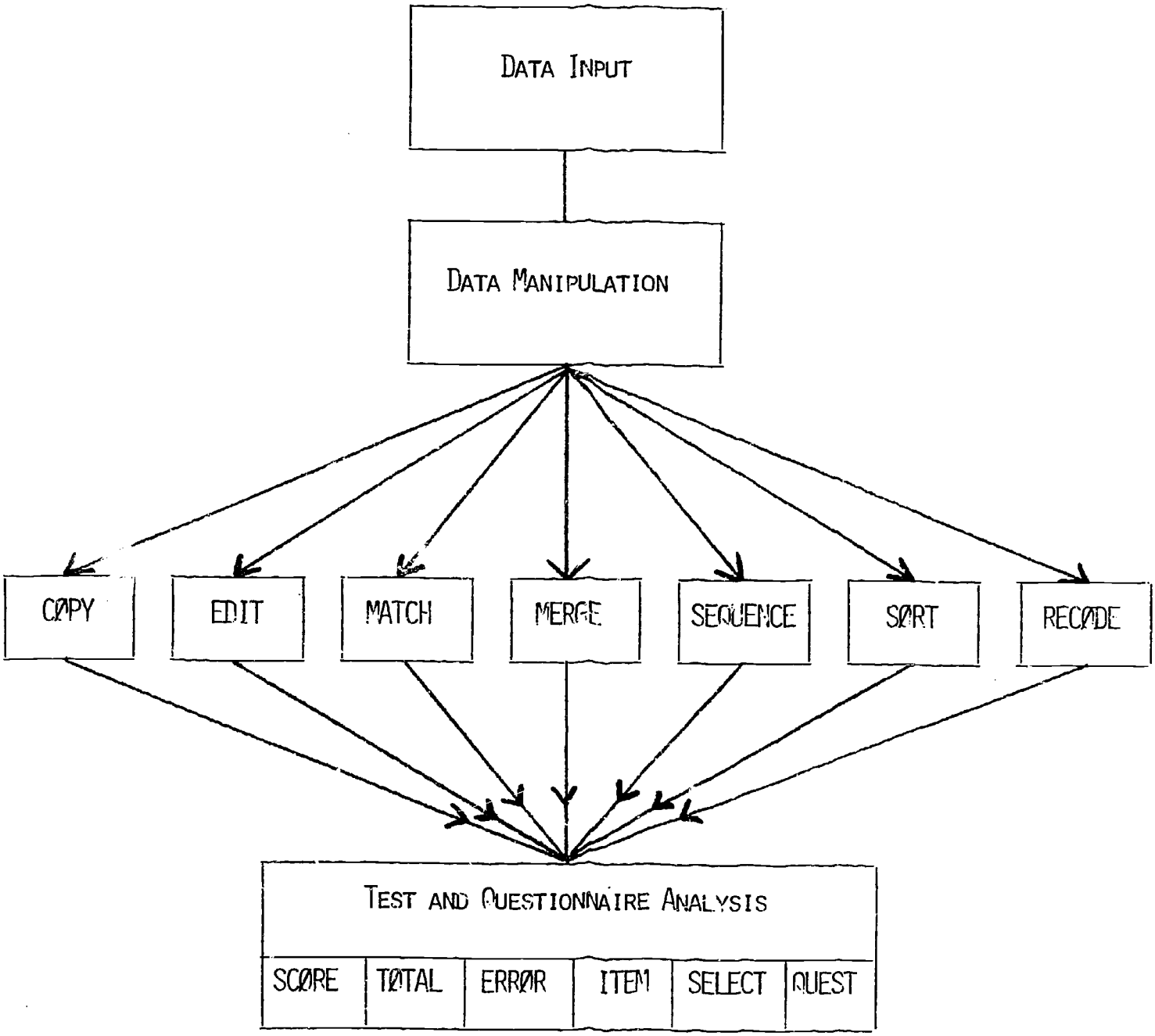


Fig. 1. The MERMAC system.

SCORE ANALYSIS OUTPUT

ID	NAME	SCORE 1	SCORE 2	TOTAL
262888459	BAMAN	4	8	22
111111111	ELMQUIST	9	7	38
326421537	EWOLDT	10	10	48
444321111	HEIKOFF	11	10	40
353420511	KOPCHEL	8	11	42
349374774	LEE	9	14	53
361448120	MCKAY	10	8	37
330323181	NORREGAARD	6	6	28
777888888	OWEN	8	12	46
357417445	OWENS	7	9	33
331448266	PEARSON	6	11	33
322407590	PL SKO	5	1	32
111110000	PLUTCHAK	6	5	28
343441821	RODELL	10	14	50
340448994	ROTHBAUM	13	11	51
999999010	SCALETTA	8	13	43
328409879	SOBOL	8	10	44
346463788	STUDKMAN	6	11	37
347400415	SUTHERLAND	6	11	37
075449011	SWICK	6	9	27
550761471	TYMCHEK	9	10	30
341360561	VALENZIA	6	9	30
327480185	WARD	10	14	46

Figure 2. Sample output from SCORE

2. Take scores for a group of individuals and produce a frequency distribution and histogram, mean, median, standard deviation, Kuder-Richardson reliability, standard error of measurement, and Spearman-Brown prophesy for a reliability of .90. In addition, individual raw scores, standard scores, and percentiles may be listed. Individual raw scores and standard scores can be weighted, summed, and the sum assigned a letter grade. All this data can be easily provided to the student. An example of the TOTAL program output is presented in Figure 3.

SUMMARY OF TEST STATISTICS

NUMBER OF ITEMS	15
MEAN SCORE	3.65
MEDIAN SCORE	3.23
STANDARD DEVIATION	2.40
RELIABILITY (KR-21)	0.558
S.E. OF MEASUREMENT	1.60
POSSIBLE LOW SCORE	0
POSSIBLE HIGH SCORE	15
OBTAINED LOW SCORE	0
OBTAINED HIGH SCORE	14
NUMBER OF SCORES	173
BLANK SCORES	0
INVALID SCORES	0
VALID SCORES	173

"SPEARMAN-BROWN PROPHESY FORMULA: IN ORDER FOR THIS TEST TO OBTAIN A RELIABILITY OF .90 IT MUST BE 6.14 LONGER. (92 ADDITIONAL ITEMS)."

TEST FREQUENCY DISTRIBUTION

RAW SCORE	STANDARD SCORE	PER CENTILE	PERCENT	FREQ	CUM FREQ	EACH # REPRESENTS 2 PERSONS
14	931	99	0.6%	1	173	#
13	889	99	1.2%	2	172	#
12	848	98	0.0%	0	170	
11	806	98	1.7%	3	170	##
10	764	97	0.6%	1	167	#
9	723	96	0.6%	1	166	#
8	681	95	0.6%	1	166	#
7	639	95	4.0%	7	164	####
6	598	91	5.2%	9	157	#####
5	556	86	10.4%	18	148	########
4	514	75	19.1%	33	130	##################
3	473	56	22.5%	39	97	#####################
2	431	34	20.8%	36	58	###################
1	389	13	10.4%	18	22	########
0	348	2	2.3%	4	4	##

TOTAL TEST PROGRAM USING SUMMING AND GRADING
SECTION 063

	PART-ONE			PART-TWO			PART-THREE			PART-FOUR			TOTAL			SUMMED STD SCORE		
	RAW	STAND	PCT	RAW	STAND	PCT	RAW	STAND	PCT	RAW	STAND	PCT	RAW	STAND	PCT	RAW	STAND	PCT
361443742ANDERBERG	3	371	23	4	475	51	2	365	18	3	405	39	12	322	7	39	328	7
GRADE IS E																		

Figure 3. Sample output from TOTAL

3. Return to each student a page containing his test-score and a list of the items he missed with his responses and the correct responses. An example of the ERRØR program output is presented in Figure 4.

STUDENT ERROR REPORT

NAME: ANDERSON — TEST SCORE: 11

*** HERE IS A SUMMARY OF THE 4 ITEM(S) YOU MISSED***

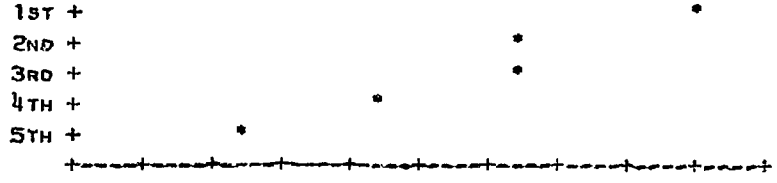
ITEM No.	YOUR RESPONSE	CORRECT RESPONSE	ITEM No.	YOUR RESPONSE	CORRECT RESPONSE	ITEM No.	YOUR RESPONSE	CORRECT RESPONSE	ITEM No.	YOUR RESPONSE	CORRECT RESPONSE
6.	C	D	7.	D	C	12.	A	C	15.	D	C

Figure 4. Sample output from ERRØR

4. Analyze his item data by providing a plot of the percentage of individuals responding to the keyed response by fifths of the total score distribution. For each item alternative the proportion of individuals responding, a point biserial correlation, and the number responding to each alternative by fifths is provided. An example of the ITEM program output is presented in Figure 5.

ITEM ANALYSIS OUTPUT

ITEM 4 PERCENT OF CORRECT RESPONSE BY FIFTHS



MATRIX OF RESPONSES BY FIFTHS D IS CORRECT RESPONSE

	A	B	C	D	E	OMIT
1ST	1	2	0	39	1	0
2ND	3	1	2	22	5	0
3RD	5	1	1	26	6	0
4TH	11	0	2	17	6	0
5TH	4	1	5	6	6	0
PROP	0.14	0.03	0.06	0.64	0.14	0.00
RPBI	-0.21	0.01	-0.18	-0.37	-0.19	0.00

Figure 5. Sample output from ITEM

5. Analyze his item data by using some external criterion rather than the keyed test score. The output format of SELECT is identical to that of ITEM.
6. Summarize item data from questionnaires or tests with no known correct answers by providing a frequency distribution of responses, a weighted mean, and a standard deviation for each item. In addition, subscores may be generated with means, standard deviation, split-half reliabilities, and percentage of individuals responding to the contributing items. It is also possible to assign deciles to the item and subscore means based on a table look-up. An example of the QUEST program output is presented in Figure 6.

QUEST ANALYSIS OUTPUT

SEX DISTRIBUTION

FEMALE	MALE	OMIT
0.13	0.63	0.24

EXPECTED GRADE

A	B	C	D	E	OMIT
0.19	0.25	0.25	0.06	0.00	0.25

ITEM	SA	A	D	SD	OMIT	BEST	MEAN	S.D.	DECL	0123456789
1.	0.25	0.44	0.06	0.13	0.13	SA	2.93	0.95	4	*
2.	0.13	0.00	0.38	0.35	0.13	SD	3.14	0.68	7	*
3.	0.06	0.50	0.19	0.13	0.13	SA	2.57	1.01	5	*
4.	0.06	0.44	0.25	0.13	0.13	SA	2.50	0.80	3	*
5.	0.56	0.25	0.06	0.00	0.13	SA	3.57	0.70	8	*
6.	0.38	0.31	0.13	0.06	0.13	SA	3.14	0.85	6	*
7.	0.44	0.25	0.06	0.00	0.25	SA	3.50	0.81	9	*
8.	0.06	0.44	0.25	0.06	0.19	SA	2.38	0.80	6	*
9.	0.19	0.44	0.13	0.06	0.19	SA	2.92	0.70	6	*
10.	0.00	0.19	0.63	0.06	0.13	SD	2.86	0.89	6	*

--SUBSCORE--	ITEMS	RESP	MEAN	S.D.	REL	N-01
TOTAL	10	0.85	2.95	0.88	0.92	6
PART-ONE	5	0.88	2.94	0.91	0.98	5
PART-TWO	5	0.82	2.95	0.85	0.90	6

SAMPLE SIZE = 16

Figure 6. Sample output from QUEST

The MERMAC system is written in Basic Assembly Language (BAL) for IBM System/360 models 40 and above which have Operating System (OS) with Queued Sequential Access Method (QSAM) support.

Application

Let us assume that the first basis of assisting the instructor in developing valid and reliable tests is also the basis for developing an evaluation model. The first step in the model, therefore, must be to provide assistance to the instructor in the determination of instructional objectives and the appropriate methods of measuring those objectives. This should also involve helping the instructor to write and select his test items. Figure 7 presents the interface of the model and the system.

The first step in the process which involves MERMAC would be to take the newly constructed or previously used items and administer them to the class and then subject them to an item analysis using either the keyed score or an external criterion score. Through the use of the item analysis statistics and the content knowledge of the instructor, a reliable and valid set of questions can be generated.

Once the final set of questions has been obtained, then the instructor is ready to have them administered, scored, and analyzed to determine what number of subscores exist or should be constructed. The instructor could now expect to receive score distributions on the total test or subscores for students in his course. These score distributions would allow him to determine how well the students were accomplishing his course objectives and what grades should be assigned. At the same time,

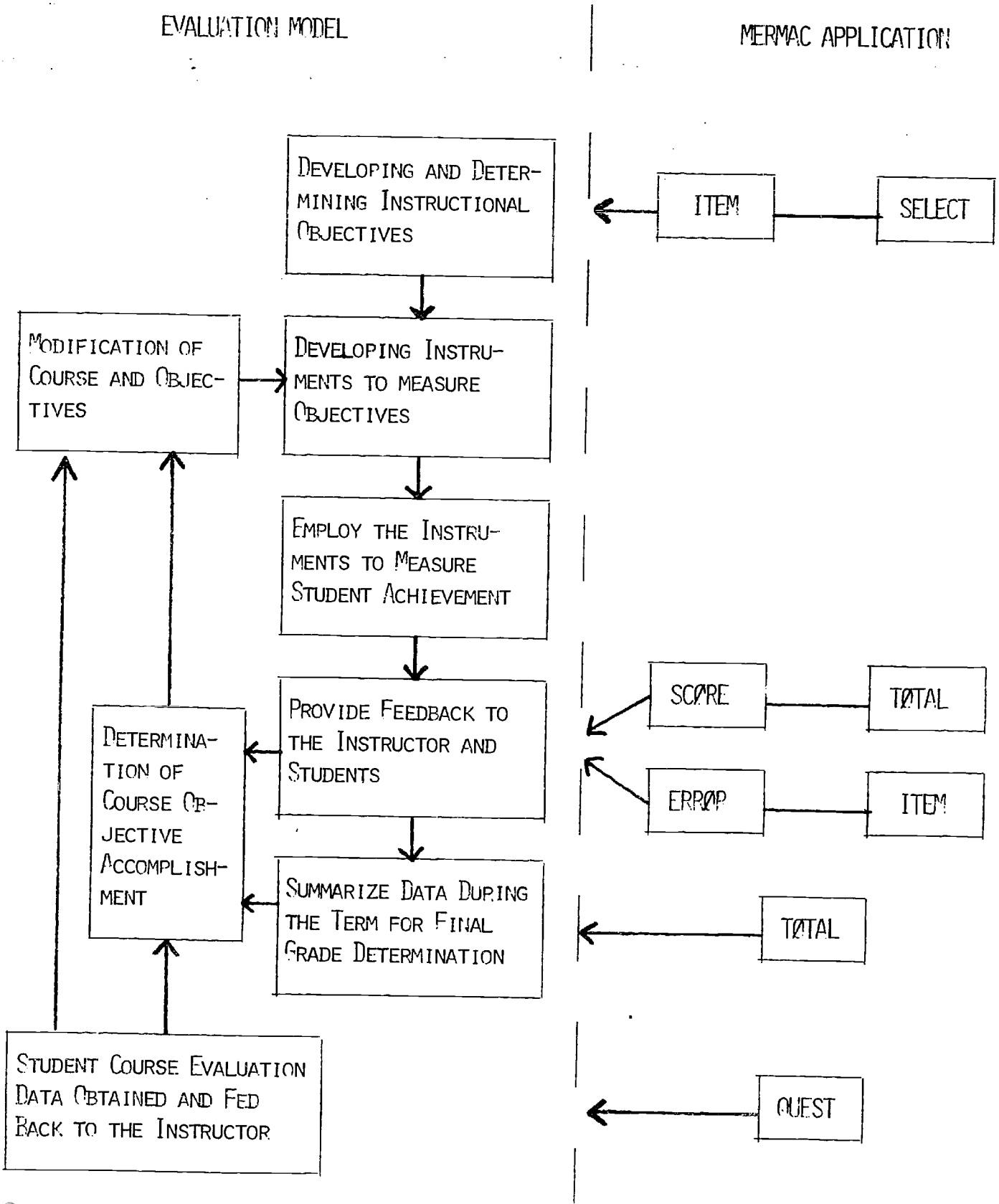


Fig. 7. Interface of the evaluation model and MERMAC.

the students can be made aware of their performance on specific questions and in comparison to their classmates through the Student Error Reports.

The instructor should discuss with the students those items most frequently missed and then provide them with a mimeographed sheet outlining the topical areas covered by the test questions showing how they fit into his scheme or course objectives. By allowing the students only to take the mimeographed sheet along with their Student Error Reports, the instructor is able to help the students identify their weakest areas and concentrate on improving them rather than memorizing the correct answer to each specific question posed. In addition, this allows the instructor to begin building a secure, reliable, and valid item bank for use later in that course or in other courses.

Once an instructor has developed the objective tests, term papers, laboratory exercises, etc., to be used in his course, he can now use the summing and grading option of the TØTAL program to accumulate and weight the scores and grades he assigns. This not only relieves the instructor of the time-consuming clerical operations involved but also gives him a more objective method of combining the scores and grades assigned. In addition, each student can be provided with his complete record of performance during the course along with his final grade.

In order to satisfy the second basis of returning results to the instructor and students in the fastest possible time, particular types of hardware and support staff are needed. The hardware should consist of a keypunch, an optical scanner, and either an IBM 360 system model 40 or above or IBM 370 system model 135 or above. The support staff could consist of only individuals capable of operating the keypunch and optical scanner, assuming that each institution has a staff available to maintain the computer operation.

The only part of the proposed model that has not been dealt with is the Questionnaire Analysis. The QUEST program provides a means of collecting student attitudes and opinions toward a course and instructor, analyzing them, comparing the results to normative data, and then returning the results to the instructor. An example of this would be the Illinois Course Evaluation Questionnaire (CEQ) (Spencer & Aleamoni, 1970) which yields responses to 50 items and six subscores (general course attitude, method of instruction, course content, interest-attention, instructor, and specific items). The subscore can then be compared to norms developed for different courses, course levels, departments, colleges, rank of instructor, etc. The instructor can also obtain results on any items that he may wish to generate.

Questionnaire data, in particular course evaluation data, can also help the instructor in diagnosing his course for possible changes in the (a) method of instruction, (b) course content, (c) personal delivery, (d) types of tests, (e) instructional materials, etc.

Summary

The MERMAC system, therefore, can provide a course evaluator and the instructor with various types of data that would be helpful in improving his course. The other essential elements needed to ensure that the data is effectively utilized in improving any course is a qualified, motivated course evaluator and an interested, willing instructor.

Reference

Spencer, R. E. & Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 7(3), 209-210.